STATISTICS WORKSHOP III

United States Department of Agriculture

Hypothesis Tests and Interval Estimation

presented by
Vicki Lancaster
NEPTUNE AND CO., INC.
vlancast@neptuneandco.com

IntervalInterval EstimationInterval Estimation & Hypothesi

Goals

" IntroduceIntroduce theIntroduce the terminology ofIntro *TypeType I error, Type II error, power, testest statitest s significance level*, and *margin of error*.

" ConductConduct *hypothesis tehypothesis testhypothesis tes intervals* for comparing for comparing a sam for c hypothesized value.

" ExploExplorExploreExplore the relationship between *hypothesis test* and a *confidence interval.*

" ExploreExplore the relationship between a *one-sided hypothesis test* and a *confidence limit*.

IntervalInterval EstimationInterval Estimation & Hypothesi

StatisticStatisticalStatistical infStatistical inference is a formal conclusionsconclusions from sample data that take into acco thethe effects of chance variation. Probability is used to quantifyquantify how confident the researcher is that qu conclconclusconclusionsconclusions drawn from the sample and not the result of a chance occurrence.

TheThe objective of an inferential method isThe objective inferencesinferences about a population binferences informationinformation contained in a sample drawn infor population.

Inductive Inference

Population

Sample Sample Sample Sample

Hypothesis Tests ***Introduction***

ThereThere are two *inferential* te tech techniques a researche useuse to draw conclusions fromuse to draw conclusions fron *hypothesis testing* and the other is *interval estimation*.

Definition . . .

" TheThe *hypothesis test* or *test of significance* is based on thethe concept of proof by contradictiothe concept c composed of four parts.

  " State the *null* and *alternative hypotheses*.

  " Collect and assess the data.

  " CalculateCalculate the *test statistic* and the probabilit ofof observing that particular value of the of observi statistic or one more extreme.

  " *Make a decision* and state the conclusions.

Hypothesis Tests *Null & Alternative Hypotheses*

TheThe purpose of a *test of significance* is to g is to giv is to statementstatement of the strength ofstatement of the strength data against the null hypothesis.

Definitions . . .

" TheThe statement being tThe statement being tested Th (abbreviated H_o). Most). Most often). Most often the nul isis a statement of the stais a statement of the status quo no difference between two orno difference between two c

" TheThe statement the researcheThe statement the resear <u>*alternativealternative hypothesis*</u> (abbreviated H_a). Th alternalternatalternativealternative hypothesis is also re *research* or *workiworking hypothesis*. The research hypothesishypothesis is usuallhypothesis is usually h collected.collected. To verify the reseacollected. To veri investigatorinvestigator tries to continvestigator tries to co in other words, a proof by contradiction.

Hypothesis Tests ***Null & Alternative Hypotheses***

The *null hypothesis* can be thought of as . . .

" aa statement of no difference between twoa statement of r populations or treatments;

" aa statement of noa statement of no difference betweena st or treatment and a hypothesized value;

" aa statement that the effect the researchea statement that th is not present; and

" aa statement that the researcher hopes to find evidence against.

Hypothesis Tests *Null & Alternative Hypotheses*

Why is a *null hypothesis* a statement of no difference?

InIn most investigations aIn most investigations a In m decisiondecision based on incomplete information, thatdec informationinformation is based on a sample not the en population.

BecBecauseBecause decisions are made based on incompl information,information, a researcher caninformation, a resea rather, a researcher can only *disprove* a claim.

ToTo see the consequences of thTo see the consequences of t
hypothetical example.

Suppose Company X manufactures marbles usingSuppose Company X manufactures marbles u
outputoutput is automatically collected in a single container. Millions of marbles
areare produced each day. Noware produced each day. Now supposeare produced each day. N
onlyonly whiteonly white monly white marbles. In order to verify that the marbles are
operatoroperator samples theoperator samples the contaioperator samples the container by p
severalseveral timseveral times thseveral times throughout the day. One day the ope
malfunctionmalfunction in the manufacturing process which results inmalfunction in the manufa
blackblack marbles, but when the operator samples the containerblack marbles, but when the
whitewhite marbles. Has he *proven* that all the marbles are white? No that all the marbl
operatoroperator is seeing only a small portion of the millions of maroperator is seeing only
conclusionconclusion can *only* depend on w depend on what depend on what he sees. On
samplesample had contained a black marble, this clearly *disproves* the claim the claim that all
the marbles are white.

Hypothesis Tests ***Null & Alternative Hypotheses***

In the preceding example the null hypothesis is

H_o: *the marbles are all white,*

a statement of the status quo.

The alternative hypothesis is

H_a: *the marbles are not all white.*

TheThe aThe alternative hypothesis is why the data are beiT collected,collected, thecollected, the operator suspects the statu incorrect.

ByBy not By not findBy not finding any nonwhite marbles, the o notnot proven the null hypothesis is true; the onot proven cancan can onlycan only can only concan only conclude that the to reject his claim.

Hypothesis Tests *Null & Alternative Hypotheses*

TheThe The above example illustrates that the *null hypothenu*
cacancan never be proven, but rather, it is either rejec, but rath
notnot rejected based on the information gathered fnot reject
an investigation.

IfIf the null hypothesis is not rejected,If the null hypothesis is n
notnot proven that it is true, but that it is true, but rather, that it
onlyonly only cononly conclude there is insufficient infonly co
the null hypothesis.

ThiThisThis is whThis is why the null hypothesis is the
researcherresearcher hopes toresearcher hopes to findresearch
collected in an attempt to reject the null hypothesis.

Hypothesis Tests *Null & Alternative Hypotheses*

It is important to remember....

> *FailingFailing to reject the nFailing to reject the nuFa considered considered proof that the null hypotconsidered proo*
> AA difference may exist, but due to an insufficientinsufficient number of experimentainsuffici inhomogeneousinhomogeneous experimental inhomoge imimpreciseimprecise measurement techniques, thi investigationinvestigation may yield dinvestigation ma null hypothesis is not rejected.

Hypothesis Tests ***Null & Alternative Hypotheses***

ThereThere are two typThere are two typesThere are two types toto in the stto in the statito in the statistical literature as *d* *directidirectiondirectional*)) and *two-tailed* (*two-sided*, *nondirec* TheThe type of *alternative hypothesis* determines determ location of the region of rejection for the H_o.

InIn a *directional alternative*, the researche, the resea informationinformation to suggest the direction of theinforma For example,

$$H_o: \quad 0 \quad \textit{versus} \quad H_a: \quad > 0.$$

TheThe research believes the population mean is greaThe res than 0.

Hypothesis Tests ***Null & Alternative Hypotheses***

InIn a *nondirectional alternative*, the researcher h, the resear
information to suggest a direction. For example,

$$H_o: \quad = 0 \quad \text{versus} \quad H_a: \quad 0.$$

TheThe researchThe researcher doeThe researcher does not
population mean will deviate from 0.

Hypothesis Tests ***Null & Alternative Hypotheses***

The *null* and *alternative hypotheses* form a dichotomy wherewhere only one of the hwhere only one of the hypothe nullnull hypothesis is rejected, than tnull hypothesis i hypothesis must be true.

ProbabProbabilitProbabilityProbability is used to quantify evidenceevidence against the null hypothesis. Probability tell thethe researcher what would happen ifthe researcher what w ggengeneratedgenerated the data were *repeated again* and *aga again*....

ThiThisThis This describes a *frequency*-based view o probabilityprobability and is the fprobability and is the f statisticsstatistics (classstatistics (classical stats tectechniquestechniques such as the analysis of variate and the *t*-test).

Hypothesis Tests *Test Statistic*

TheThe evidence against the null hypothesis is quantified bbyby calculating a *test statistic* and determinin and dete probabiprobabilprobabilityprobability (commonly referred t observingobserving that particular result orobserving that pa

Definition . . .

" InIn general, the decision to reject or noIn general, the de nullnull hypothesis is based on a single value that summarizessummarizes the data. summarizes the data calledcalled thecalled the *test statistic*. The test statistic is. T *variable*,, associated with a test sta, associated with a probabilityprobability (commonlyprobability (commonly ofof observing that particof observing that particularof extreme.

Definitions . . .

" AA *random variable* assig assigns assigns an event a num thethe realthe real line. For example,the real line. For exar gamesgames a team will win in a given seasongames random variable.

" TheThe mechanism used to asThe mechanism used to a randrandomrandom varandom variable are *probability di probabilityprobability probability distribution* of of of a of a of a aa value of a random variable a number in the intervalinterval [0, 1] for a given event. The probability ddistrdistributiondistribution is represented by a fun exampleexample thexample theexample the Student s probability distribution function.

Hypothesis Tests ***Test Statistic***

For example . . .

SupposeSuppose there is an urn wSuppose there is an urn w areare black and 2are black and 2 of which are red. In other w population is {B1, B2, B3, R1, R2}.

WeWe are going to employ simple random We are going to selectselect samples ofselect samples of size 3. There are 10 diff that we can construct.

LetLet *Q* represent the represent the number of red balls in th of 3 balls. *Q* is the random variable.

What are the possible values *Q* can take on?

Population {B1, B2, B3, R1, R2}

Event selecting 3 balls using simple random sampling without replacement **Set** of possible **outcomes**	B1, B2, B3	B1, B2, R1 B1, B2, R2 B1, B3, R1 B1, B3, R2 B2, B3, R1 B2, B3, R2	B1, R1, R2 B2, R1, R2 B3, R1, R2
Values of the **Random Variable** q = the # of red balls	0	1	2
Probability of observing that value of the **Random Variable**	1/10	6/10	3/10

Hypothesis Tests *Test Statistic*

We have three pieces of information here,

1. a random variable,

2. all possible values of the random variable,

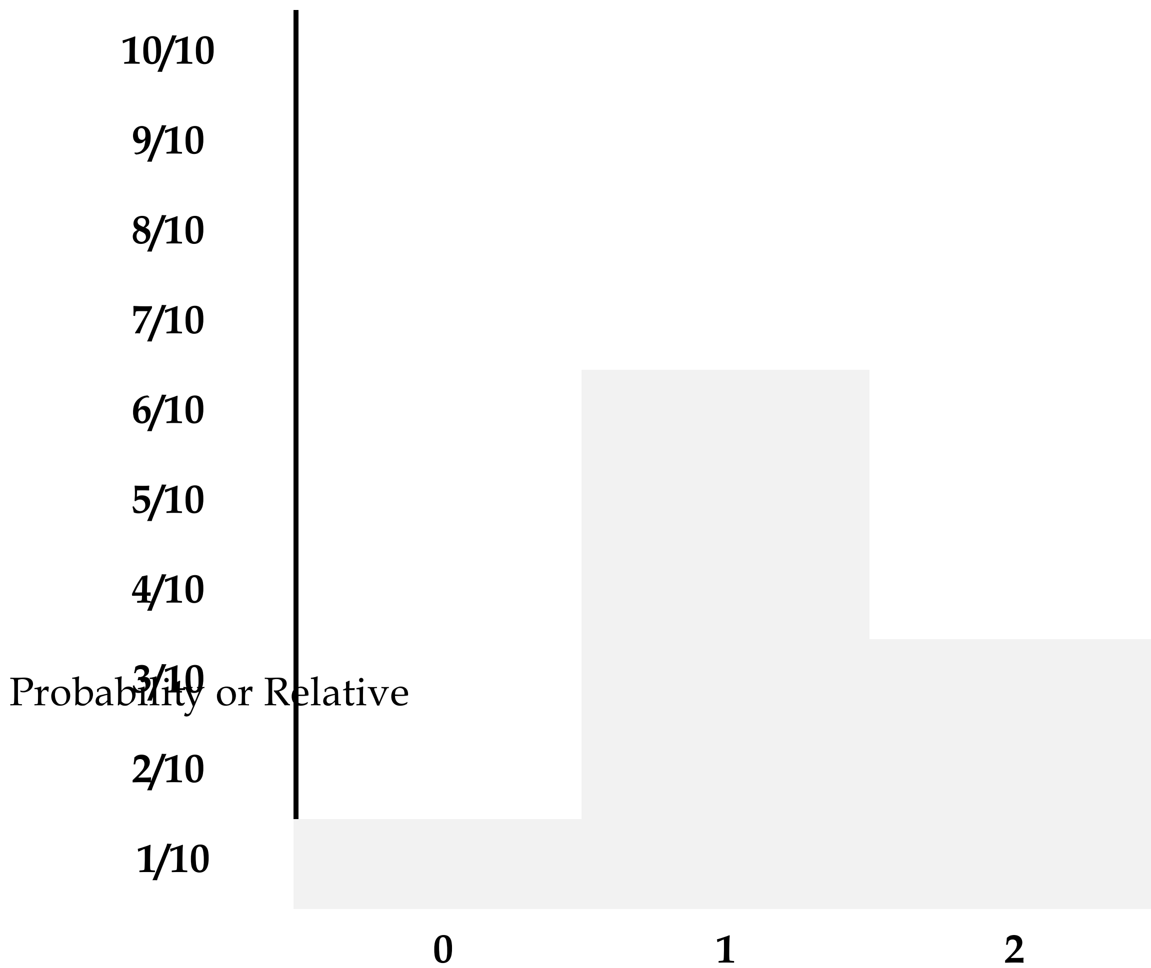
3.3. the probability of occurrencethe probability of occurrence

TheseThese are the three pieces of informationThese are the constructconstruct a *probability distribution* for a randon variable.

Probability Distribution for the Variable *Q*

q	0	1	2
P(Q = q)	1/10	6/10	3/10

A histogram or
probability distribution function (PDF)
of the random variable on page 17.



Hypothesis Tests *Test Statistic*

Interpretation: of the PDF.

" TheThe probability of observing 0 red balls when selecting a sample of 3 balls is

$$P(Q = 0) = 0.1.$$

" TheThe probabilitThe probabilityThe probability of ob selecting a sample of 3 balls is

$$P(Q = 1) = 0.6.$$

The PDF is a *relative frequency histogram*

Definitions . . .

" *Frequency* is the number is the number o is the nur particularparticular class.particular class. The *relativerelat* expressedexpressed as a expressed as a prexpressed as a frequency.

Hypothesis Tests ***Test Statistic***

ThereThere is an alternative way we can express the probabilityprobability information about the randomprobabil

The *probability distribution* answers the question,

> Wha What What is What is the probability that the *equals* a specific value?

SometimesSometimes we are interestedSometimes we are int question.

For example,

> What What is What is the probability that the variable *Q isis or equal* to a specific value?

Hypothesis Tests *Test Statistic*

GivenGiven the original *probability distribution*, the question,

What What is What is the probability What is the probabili
equal to a specific value? , is easy to answer.

Probability Distribution for the Variable *Q*

q	0	1	2
P(Q = q)	1/10	6/10	3/10

Answer to the New Question for the Variable *Q*

q	0	1	2
P(Q q)	1/10	7/10	10/10

TheThe answer to this newThe answer to this new question ab
the probabilities that precede the value of interest.

Hypothesis Tests ***Test Statistic***

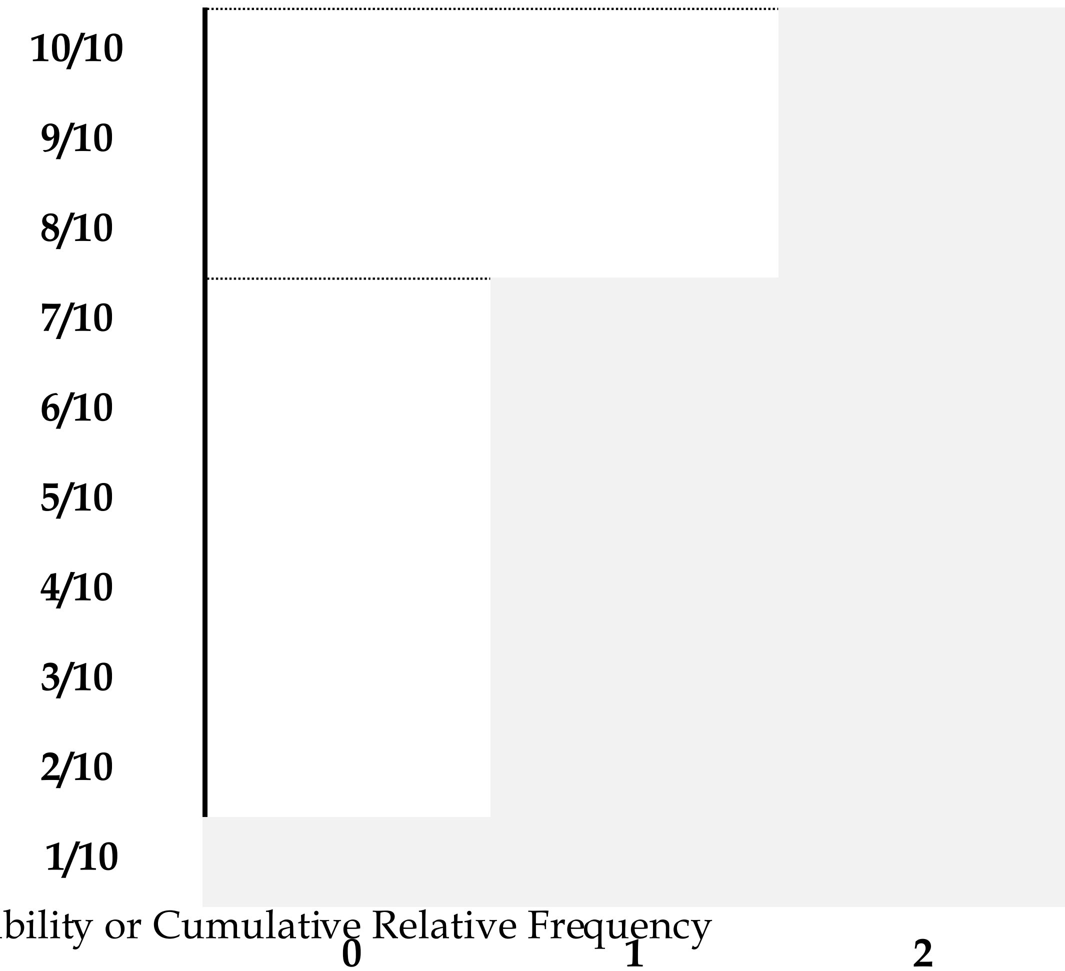
TheThe tableThe table thaThe table that gives the answers to
for the variable Q has a formal name.

It is called a *Cumulative Distribution Function* (CDF).

CDF for the Variable *Q*

q	0	1	2
P(Q q)	1/10	7/10	10/10

TheThe CDFThe CDF is another wayThe CDF is another way t aa raa random varia random variable. The CDF can be thou *cumulative relative frequency* histogram.



Interpretation: of the CDF.

" ThTheThe probaThe probability of observing 1 or fewer when selecting a sample of size 3 is

$$P(Q \quad 1) = 0.7.$$

" TheThe probability of obsThe probability of observiThe when selecting a sample of size 3 is

$$1 \quad P(Q \quad 0) = 1 \quad 1/10 = 0.9.$$

The CDF is a *relative cumulative frequency histogram*.

Definitions . . .

" TheThe cumulative frequency is the number of observatioobservationobservation leobservation less thar TheThe <u>*relative cumulative relative cumulative frelative*</u> frequencyfrequency expressed as a proportion or percent the total frequency.

What makes a *test statistic*
different from other statistics?

A *test statistic* is a *decision maker*.

ItIt has a probability distIt has a probability distribIt has a withwith it, whereas other statistics such as the mean and standardstandard deviation do not. *Test statistics* are specific a tthethe hypothesis being evaluated. For examplthe hyp commonlycommonly usedcommonly used test statistic, the *t*- evevalevaluateevaluate whether a sample mean is different hypothesizedhypothesized mean or whehypothesized mea come from the same population.

FigureFigure 1 on page 27, shows Figure 1 on page 27, show the population, sample statistic, and test statistic.

Figure 1. Distributions of
Samples Means and *t*-Test Statistics

FigureFigure 1 shows in order to make inferences,Figure 1 sh
use probability in decision making,

" thethe characteristic of interest is first summarized using a statistic, in this case the mean;

" thethe statistic is then transforthe statistic is then transfo
where

" thethe the tethe test statistic follows a known proba
distribution.

TheThe tableThe table on page 29 lists someThe table on page
test statistics and the null hypotheses they evaluate.

Hypothesis Tests *Test Statistic*

Statistical Test	Null Hypothesis	Test Statistic	Assumptions
Z-test	population mean H_o: $= {}_o$		" population variance, ^{2}, is known " population distribution is normal
Z-test	two population means H_o: ${}_1 = {}_2$		" population variances, , are known " population distributions are normal
t-test	population mean H_o: $= {}_o$		" population variance, ^{2}, is *not* known " population distribution is normal
t-test	two population means H_o: ${}_1 = {}_2$		" population variances, , are *not* known " population distributions are normal " sample variances are equal,
t-test	paired comparison H_o: ${}_d = 0$		" paired observation (2 observations on the same EU or an RBD with 2 EUs / block) " population distribution is normal
^{2}-test	population variance H_o: $^2 =$		" population variance, ^{2}, is *not* known " population distribution is normal
F-test	two population variances H_o:		" population variances, , are *not* known " population distributions are normal

One statistic is the *t*-statistic,

$$\frac{\bar{\ }}{\sqrt{\ }}. \qquad (1)$$

ItIt is a measure of It is a measure of the It is a me deviationdeviation of a sample mean from a hypothesize mean.

A theorem states that,

> ifif all possible samples of size *n* are are draw are dr fromfrom a normal populationfrom a normal population w and if for each samp and if for each sample the calculated,calculated, the frequencycalculated, the frequer valuesvalues follows the *Student sStudent s t-distribution* w (n 1) degre 1) degrees of freedom, where (n 1) is the number of degrees of freedom of .

IfIf the researcheIf the researcher iIf the researcher is wi assassumptionsassumptions about the data, then the t providesprovides the mechanism for the researcher to make inferences with *only one sample from the population*.

UsingUsing equation (1),Using equation (1), the sample mean i thatthat it follows a known probabithat it follows a known p distributidistributiodistributiondistribution of the transfo distrdistributdistribution.distribution. The PDF for the *t*-sta using the function

$$\overline{\qquad\qquad} \; \overline{\qquad} \; \overline{\qquad\qquad\qquad}$$
$$\sqrt{\quad}$$

where k = *degrees of freedom*.

A histogram of 1,000 *t*-values
with 24 degrees of freedom (*df*).

ItIt is the relIt is the relatIt is the relative frequency, distinguishes one *t*-distribution from another.

TheThe relative frequencies of the *t*-d-distri-distribution variousvarious values of the *df* are provided are provided in lineline of the *t*-table represents a particular value of the *df*.

TheThe The probability of observing a particular *t*-value i-
usedused toused to makeused to make a decision. But unfort
aa decision is made based on *incomplete information*,
mistakesmistakes can bemistakes can be made. In hypothesis
two potential decisions:

to reject or *fail to reject* the null hypothesis.

EachEach decision bringsEach decision brings with it the possib
been made.

" AA *type I error* is made if the null is made if the null is rejec
is in fact true.

" AA *type II error* is made if the null is not rejected
when it is in fact false.

TheThe table of theThe table of the next page illustrates the erro
be made in hypothesis testing.

Potential Errors in Hypothesis Testing

Make the DECISION:	The NULL HYPOTHESIS is:	
	True	False
Not to Reject the Null Hypothesis	*Correct Decision* (1)	*Incorrect Decision* Type II Error ( )
to Reject the Null Hypothesis	*Incorrect Decision* Type I Error ( )	*Correct Decision* Power (1)

Definition . . .

TheThe decision to rThe decision to rejThe decision to rej
isis in fact true isis in fact true is referred to as theis in fact
typetype of decision error can be made only if the null
hypothesis is in fact true.

IfIf the null hypothesis is inIf the null hypothesis is in factIf
ofof maof makingof making a type I error is denoted by
symsymbolsymbol .. This is commonly referred to as
statistical significance level.

BasedBased on Based on theBased on the probability of o valuevalue of the tesvalue of the test stativalue of th statistically significant or not statistically significant.

Definition . . .

> ThisThis probability ofThis probability of observing a part the test statistic is referred to as the *p-value*.

Traditionally, if the

$$p\text{-value} < \textit{type I error} \;(\quad),$$

tthethe result is labeled as *statistically significant*, and, and th nullnull null hyponull hypothesis is rejected. A *statistically* resultresult iresult is oneresult is one that is unlikely to l chance.

Definitions . . .

TheThe dThe decision not to reject the null hypothes whenwhen when it is when it is in fact false, is referred t <u>*error*</u>.. This type of decision error can. This type of decisior if the null hypothesis is in fact false.

IfIf the nullIf the null hypothesis is in factIf the null hypoth *ofof making a type II error* is denoted by the Greek symbol .

TheThe <u>*power*</u> of a of a test, 1 , is the probability o *rejecting* the null hypothesis when it is in fact false.

TheThe relationship betweThe relationship between *type I a*
illustrated in Figures 2 and 3.

Figure 2. Rejection / Acceptance Regions for the Hypotheses H_o: = 10 versus H_a: > 10 (= 0.10)

In this scenario, the null hypothesis is

$$H_o: \quad = 10.$$

TheThe variability in the sample mean diThe variabilit illustratesillustrates the fact that each time a sample is selecte from thefrom the population afrom the population a different andand therefore a different value of the sand therefore a dif calculated.

TheThe sampThe sample mThe sample mean is based on *in* thereforetherefore each timetherefore each time a sample is con calculated, the value of the mean will vary.

Figure 3. Rejection / Acceptance Regions for the Hypotheses H_o: = 10 versus H_a: > 10 (= 0.01)

In Figures 3 and 4:

" TheThe area underThe area under the *null distribution* (soli

" TTheThe area to the right of the vertical linThe area enclosed by the null distribution is .

" The area underThe area under theThe area under the *alte* line) is 1.

" TheThe area to the left of the vertical line and enclosed by the alternative distribution is .

Going from Figures 2 to 3 notice that:

" TheThe probability of makiThe probability of making T
(0.10(0.10 0.01) as the probability 0.01) as the probabilit
error increases (0.008 0.0134).

" ThisThis illustrates the *inverseinverse relationship betweenin*
II and type II errors. As. As long as . As long as the
remainsremains constant, aremains constant, a decrere
simultaneous increase in and vice versa.

SignificaSignificant in the statistical sense does not mean *important*.. It simply means not. It simply means not li chance. chance. Whachance. Whatchance. What is impo *difference*.

Definition . . .

TheThe *practical significant difference* quantif quantifie evidenceevidence against theevidence against the nullevic expertexpert judgement of the researcher rexpert judge using a *p-value*.

AA practically significant results is A practically signific differencedifference betweendifference between populatic influence the actions of the researcher.

Hypothesis Tests *Example Two-Sided*

Let s see how this works with an example.

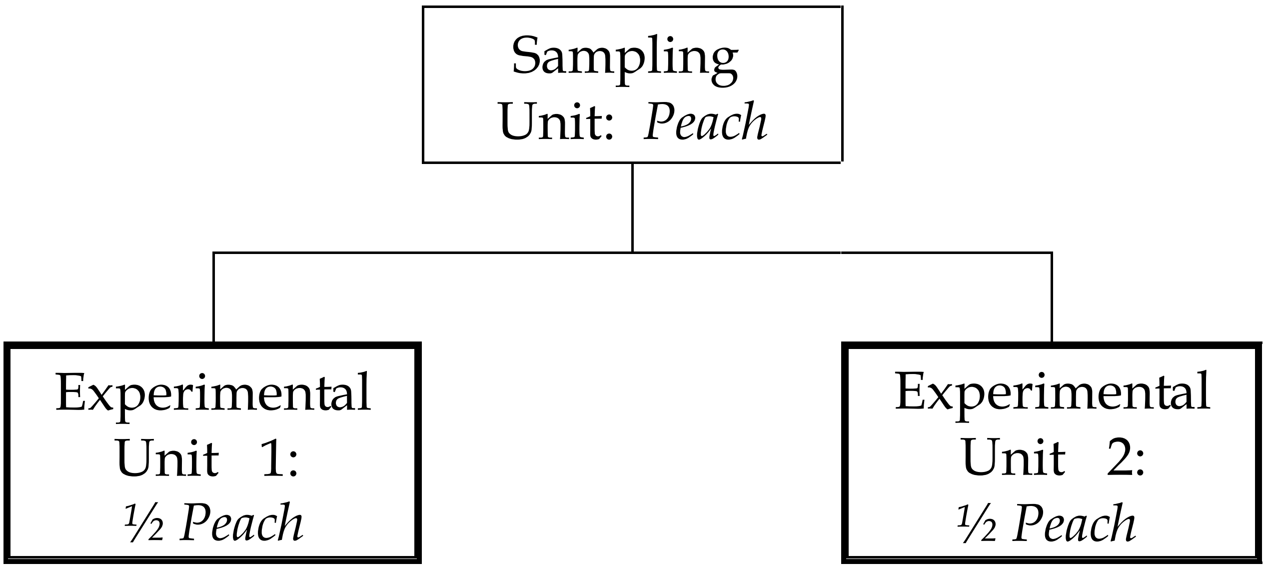
SaySay a researcher is interested in how the sugar content in peaches is affected by different storage conditions.

" PeachesPeaches of the same variety and age are randomly selected from an orchard.

" EachEach peach is cut in half and randomly assigned to one of two storage treatments.

" AfterAfter five days each peach haAfter five days each sugar content.

Since each *sampling unit* receives *both* treatments the data are paired.



1st treatment *2nd treatment*

after 5 days

measurement of EU 1

measurement of EU 2

Let,

x_{i1} = the peach half from the i^{th} peach receiving peach recei
first storage treatment, and

x_{i2} = thethe peach halfthe peach half from thethe peach hal
second storage treatment.

OneOne way to compare the relative effectiveneOne way to
twotwo storage treatmentwo storage treatments is ttwo stor
sugar content within a peach.

Let,

$_D$ = thethe population methe population meanthe pop
differencesdifferences between storage differences be
particular peach.

Let,

D_i = represrepresentrepresent the differencreprese cconcentrationconcentration concentration betweenco i^{th} peach.

AssumingAssuming n = 10 sample units were take symbolically we have,

$$D_1 = x_{11} \quad x_{12}$$

$$D_2 = x_{21} \quad x_{22}$$

$$\vdots$$

$$D_n = x_{n1} \quad x_{n2}.$$

OneOne wOne way to coOne way to compare the two treatn thesethese differences and see whether the avethese difference is greater than or less than zero.

Our estimator for the population mean, $_{D,}$ is,

— — ,

the sample mean of our *n* differences. The difference datadata aredata are used to calculate the statistic and thisdata isis transformed into a known probability distribution.

Table 1. Sugar Concentration in Peaches.

Peach	*1st Treatment*	*2nd Treatment*	*Differences*
1	$x_{1,1}$=23.69	$x_{1,2}$=30.25	D_1= 6.56
2	$x_{2,1}$=42.61	$x_{2,2}$=26.07	D_2= 16.54
3	$x_{3,1}$=30.73	$x_{3,2}$=39.25	D_3= 8.52
4	$x_{4,1}$=26.73	$x_{4,2}$=19.26	D_4= 7.47
5	$x_{5,1}$=34.84	$x_{5,2}$=35.22	D_5= 0.38
6	$x_{6,1}$=38.72	$x_{6,2}$=34.14	D_6= 4.58
7	$x_{7,1}$=45.29	$x_{7,2}$=33.43	D_7= 11.86
8	$x_{8,1}$=26.87	$x_{8,2}$=30.36	D_8= 3.49
9	$x_{9,1}$=26.24	$x_{9,2}$=22.13	D_9= 4.11
10	$x_{10,1}$=37.64	$x_{10,2}$=19.37	D_{10}= 18.27

SinceSince there is noSince there is no information to suggest o willwill pwill performwill perform better than the other, thi *hypothesis test*. The hypotheses are:

$$H_o: \quad _D = \quad _{Do} \quad \textit{versus} \quad H_a: \quad _D \quad\quad _{Do},$$

where,

$_D$ = thethe populthe populatthe population mean of t differencesdifferences differences bedifferences betw particular peach, and

$_{Do}$ = the hypothesized populationthe hypothesized popula null hypothesis, which for this scenario is 0.

ThingsThings to think aboutThings to think about are how mu
the H_o will we insist on?

ThatThat is, what That is, what is thThat is, what is the *sign*
committing a *type I error*?

For an = 0.10,

> wewe we are rewe are requiring that the data give
> againstagainst H_o so strong that it so strong that it would h
> than 10% of the time (2than 10% of the time (2 times intha
> fact true.

For an = 0.01,

> wewe are insisting on stronger evidence against H_o,
> evidenceevidence so strong that it would happen no more
> thanthan 1% of the time (1 time in 100) wthan 1% of the
> fact true.

SinceSince we are testing aSince we are testing a hypothSinc
population where,

- the *variance is unknown*, and

- the sample size is *small* (< 30),

the *t-test* statistic is appropriate,

$$\sqrt{}\ \frac{}{}\ .$$

Hypothesis Tests *Example Two-Sided*

MostMost clasMost classicaMost classical procedures are b *aboutabout the characteristics* of of the random variable, of the ra validityvalidity of thevalidity of the analyses dependsvalidity *assumptions*.. For the paired. For the paired *t*-test-test the assu

" D_i s are the result of paired measurements,

" D_i s are independent, and

" D_i s are normally distributed,

" wherewhere the assumption ofwhere the assumption of are no *outliers* .

Hypothesis Tests ***Example Two-Sided***

TheThe *graphical methods* of EDA of EDA pr of EDA
diagnosticdiagnostic tools fordiagnostic tools for confirming a
assumptionsassumptions are not metassumptions are not r
actions.

GraphicalGraphical displays meet the need
toto seeto see the behavior ofto see the behavior of the data, to
rrevrevealreveal the unexpected features,
suchsuch as *outliers*; and confirm or
disprovedisprove *assumptions*, such as the
distributionaldistributional assumptions of
normality.

TwoTwo types of plots commonly utilized in EDA are, histogramshistograms and boxplots. From thistograms and l samplesample differences, the data appear to bsa approximately normally distributed.

FoForFor coFor comparison, consider histograms of rar samplessamples (n =10) from the stansamples (n =10) from the s

ThisThis illustrates that random samples from This illustrat
distributiondistribution can look very *abnormal* when the sam
size is *small.*

TheThe boxpThe boxplotThe boxplot for the sample dif provideprovide evidence against the assumption the data are normally distributed.

ForFor comparison, let s look at the boxpFor comparison, let used to construct the histograms on page 56.

Again,Again, Again, randomAgain, random samples from a no
appearappear somewhat *non-normal* when when the sample
is *small*.

Why should we be concer should we be concerned a should
outliers ?

By looking at the test statistic,

$$\sqrt{} \frac{}{},$$

wewe can seewe can see thatwe can see that the behavior of the
extent,extent, driven by the *sample sample msample mean a*
standard deviation.

SinceSince they are not resistant measures of
locationlocation andlocation and shlocation and shape. The
andand *sample standard deviatiosample standard deviation* are b
impactedimpacted byimpacted by the presenceimpacted by th

Hypothesis Tests *Example Two-Sided*

What are *outliers* . . .

" A n An *outlier* i in any in any graph of data is an indiv observationobservation thatobservation that fallsobservati the graph.

(David Moore, *Statistics Concepts and Controversies*)

" Highly suspect observations in a sample.

(B. J. Winer, *Statistical Principles in Experimental Design*)

" . . . . observed values which . . . observed values remoteremote from the main body ofremote from the mai bebe discarbe discardebe discarded as being erroneous miscalculations, etc.

(Rupert Miller, *Beyond ANOVA, Basics of Applied Statistics*)

TheThe impactThe impact of *outliers* on $\overline{}$ and s_D depends o
the *sample size* and the *variability*.

IfIf the sample size aIf the sample size and theIf the sample
enough enough theenough then enough then the impact
value of the *t*-statistic toward the *outlier.*

In this case,

oneone is more likely *to reject* the H_o due t due to t
inflated value of the test statistic.

IfIf the sample size andIf the sample size and the varIf the thenthen *outliers* can actually have the impact can actually ha the *t*-statistic toward the hypothesized value.

In this case,

one is *less likely to reject* the H_o due to the deflated value of the test statistic.

Recommendation

EvenEven wEven when there is no assignable cause fo thethe outlierthe outlier and it is a true outlier, it doesthe necessarilynecessarily mean the datnecessarily mean the andand should be changed or omitted. Prand should b allall analyses with and without the outlier(s). WhenWhen there are no differences between the twotwo sets of analyses the dilemma is resolved. WhenWhen there is a difference, an interpretation ofof botof both of both results should be provided in conclusions.

BasedBased on the histogram (page 57) and boxplotboxplot (pboxplot (page 59boxplot (page 59) there a concernsconcerns about the assumpticoncerns about normality,normality, therefore we cannormality, therefore we the *t*-test.

Note:

ThereThere areThere are many *qualitative* (EDA) (EDA) ar (CDA)(CDA) statistica(CDA) statistical techni(CDA distrdistribdistributionaldistributional assumptions and outliers.outliers. This is the topic of Workshopout Exploratory and Confirmatory Data Analysis.

TheThe D_i values in Table 1 (page 50) are summarized below,below, this information is usebelow, this information is u

—	s_D	Do	n
4.39	9.29	0	10

The observed value of the test statistic is

$$t_{\text{obs}} = \frac{\sqrt{}}{} = 1.49.$$

ThTheThe next step is to find the *critical values* for a for a t sidedsided test basedsided test based on the t-distribution wi = 0.10, $\pm\, t_{(9,\ 0.10/2)}$.

Hypothesis Tests ***Example Two-Sided***

TheThe *critical values* partition partition the quantiles partitio intointo two regions, the region were thinto two regions, th and the region where the H_o is not rejected.

SinceSince t_{oobobs} = 1.49 is less then the absolute value of the criticalcritical values $|\, t_{(9,(9,\, 0.10/2)} \,|$ = 1.83, we | = 1.83, we fail to re hypothesis.

TheThe *critical values* can be foun can be found usin can be fo containscontains the quantiles of thecontains the quantiles of th and values.

Example of a *t*-Table

df	=0.1*	=0.05	=0.025	=0.01	=0.005	=0.001
...	...	...	...	...	...	...
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.405	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	**1.833**	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
...	...	...	...	...	...	...

*The area under the curve to the right of the *t*-value (quantile).

t-tables-tables are usually found in th-tables are usuall statistical methods books.

Hypothesis Tests *Example Two-Sided*

TheThe hypotheses cThe hypotheses caThe hypotheses can a
the *p*-value for t_{obs}.

" AA *p*-value can be thought of as an value can be t
significance level.

" CalculatingCalculating aCalculating a *p*-valuevalue gives
thanthan simply compthan simply comparing than sim
test statistic to a *critical value(s)*.

TheThe *p*-value is thevalue is the probability of observing avalu
value of t_{obs} or one more extreme.

ForFor theFor the peacFor the peach investigation, the *p-va* probabilityprobability of obseprobability of obserprobability 4.394.39 or less than 4.34.39 when th4.39 when the true population differences is 0.

SinceSince 4.39 corresponds to aSince 4.39 corresponds to a *t*- needneed to calculate the probability of observing this resultresult or one larger,result or one larger, given tresult o true.

MostMost often the *p*-value-value is calculated-value is calcul pacpackagepackage tpackage that calculates the probabi distribution.

Hypothesis Tests *Example Two-Sided*

ForFor a $t_{obs} = 1.49$, the area under the curve to the = 1.49, the ar
the value is 0.085.

SinceSince ourSince our alternative hypothesis is *two-sided*,, w nownow now conow consider results that are equally extre respect to the other side of the null hypothesis.

StatedStated anotherStated another way, this means weStated sample means that correspond to a t_{obs} 1.49.

TheThe probability of observing this setThe probability of
moremore extreme with respect to the null hypothesis is
given by,

$$\textit{the observed } p\text{-value} =$$

$$\text{Prob}(t_9 \quad t_{obs}) + \text{Prob}(t_9 \quad t_{obs}) =$$

$$\text{Prob}(t_9 \quad 1.49) + \text{Prob}(t_9 \quad 1.49) =$$

$$0.17$$

SinceSince 0.17Since 0.17 is greater than our *significancesignifica*
concludeconclude there is not enough evidence toconclude the

BothBoth approaches lead to the same coBoth approaches difference is:

" perperformiperformingperforming the test using a *critic* usus whether or not we sus whether or not we sl whereas

" thethe *p-value* quantifies the quantifies the weight of evid a particular sample provides against the H_o.

Hypothesis Tests *Example One-Sided*

Now let s change the setup a bit.

SupposeSuppose we want to test a new storageSuppose w *against* a standard one currently a standard one currently in a st isis not willing to employ the new is not willing to employ performs as well as or better than the current one.

ThisThis information is usThis information is used to This in aa *one-tailed hypothesis*. . If we let storage treatment 2 be the standard treatment, where

$$D_i = x_{i1} \quad x_{i2},$$

then hypotheses are not,

$$H_o: \quad {}_{D} \qquad {}_{Do} \ \textit{versus} \ H_a: \quad {}_{D} > \quad {}_{Do}.$$

Hypothesis Tests *Example One-Sided*

ForFor a less thFor a less than 10For a less than 10% c
hypothesishypothesis when it is inhypothesis when it is in fact
value for this *one-sided* test is $t_{(9, 1 \quad 0.10)}$ = 1.38.

TheThe difference between the *one-sided* test and the *two-sided* test is that wit test is that with test is that with the *one*
only one direction are considered important.

InIn a one-sided test the entire significance levelIn a one-sided
side of the distribution.

ItIt isIt is easier toIt is easier to see a sta
signsignifsignificantsignificant differ
one-sided test as opposed to a
two-two-sidtwo-sided test since we
onlyonly looking in only looking in
oofof the alternative hypothesiof the

Hypothesis Tests *Example One-Sided*

ForFor the one-sided test, the obsFor the one-sided test, the statisticstatistic is still 1.49. Therefore,statistic is still 1.49. Th = 0. = 0.10 *significance level* since $t_{obs} > t_{(9,\ 1\ \ 0.10)}$ = 1.3 = 1. where 1.38 is the critical value.

WhatWhat is the observed p-value-value that corresponds to-v
1.49?

TheThe observed p-value = 0.085. Since the observed p-valuevalue is smaller than the *signsignificasignificance level,* concludeconclude thereconclude there is sufficient evidenceco
the 0.10 level.

AllAll other things being equal, a *one-sided test isone-sided t* *powerfulpowerful than a two-sided test*, when the, when the rese informationinformation to specifyinformation to specify the d

ByBy more *powerful*, it is meant there i, it is meant th probabilityprobability of rejecting the null when in is in fact fal

TheThe *confidenceconfidence interval* is the other *inferential* techr
aa researcher can usa researcher can use to da researcher ca
data.

Definitions . . .

" TheThe *confidence interval* includes includes a point includ
thethe population parameter (fothe population parameter (
mean)mean) accompanied by mean) accompanied by a
associatedassociated with theassociated with the poin
error).

" TheThe two extreme points in a confiThe two extreme p
areare thare theare the lower and upper *confidence*
conficonfidencconfidenceconfidence bounds. The *confidence li*
rangrangerange of varange of values within which ther
levellevel of level of confidence that the true popu
parameter will fall.

LikeLike the alternative hypothesis, thereLike the alternative l
confidenceconfidence intervals, referred to inconfidence i
literatureliterature as *one-tailed* (*one-sidedone-sided* or *directional*
tailed (*two-sided* or *nondirectional*).

AA *two-sided confidence interval* for a po for a pop
parameterparameter consists of the sample estimate for p
parameter plus and minus the *margin of error*,

{sample estimate ± margin of error}

AA *one-sided confidence limit* for a population parameter
cconsistconsistsconsists of the sample estimate minus the *m*
error or the sample estimate plus the *margin of error*.

lower confidence limit (LCL)
= {sample estimate margin of error}, or

upper confidence limit (UCL)
= {sample estimate + margin of error}.

Definition . . .

" TheThe measure of confidenceThe measure of confiden intervalinterval includes theinterval includes the true po isis called the <u>*confidence coefficonfidence coefficient*</u>. proportionproportion ofproportion of all possible sample yieldyieldingyielding confidence intervals that contain true true population parameter. The confid true coefficientcoefficient iscoefficient is represented symbolic

IntervalInterval estimationInterval estimation and hypothesis related,related, in fact, a confidence interval can be used to conduct a hypothesis test.

ForFor example, to test H_o: = $_o$ versus H_a: $_o$ usingusing interval estimation, theusing interval estimation, t calculated from the sample data and

ifif the interval contains $_0$, , the, the hypothesized value,

thethe *null hypothenull hypothesisnull hypothesis is acc* rejected.

AA (1)100% two-sided confidence interval for the meanmean is equivalent tmean is equivalent to a twmean is meanmean atmean at the level of significance. This relatio isis illustrated in Figure 4 using the hypotheses H_o:: : = 0 versus H_a: 0.

Figure 4. Hypothesis Testing *versus* Interval Estimation: Rejection / Acceptance Regions for the Hypotheses H_o: = 0 *versus* H_a: 0, (= 0.05)

ForFor thisFor this hypotheticalFor this hypothetical illustrati
tenten were randomly selected fromten were randomly selecte
withwith mean 0 and variance 2;with mean 0 and varia
samplessamples,samples, the *t*-test statistic used to con
hypothesishypothesis test and the 95%hypothesis test and the
calculated.

The null hypothesis is not rejected when

" thethe line segment crosses the solidthe line segment cross passes through the hypothesized value zero, or

" whenwhen the filledwhen the filled circle is within the fai regionregion bounded by the solid vertical lines region bou ± 2.23.

For each of the 50 samples,

" *thethe conclusions drawn based on eiththe conclusions drawn test or the confidence interval are identical*.

In only two cases did the sample data produce:

aa value of the test statistic a value of the test statistic c region; and

aa confidence interval that did not contain the hypothesized mean.

AA common ***misimisinmisinterpretation*** of the confide
interval is that . . .

> aa 90% *confidenceconfidence interval* implies there is a implie
> chancechance that the chance that the trchance that the
> lies within the interval.

ForFor a (1For a (1 0.10) 0.10) *confidence coefficient*, the probab
intervalinterval *will* contain the unknown true contain the
parameter is 0.90.

TheThe probability, 0.90,The probability, 0.90, is a measure of
thethe sample is drawn, that the the sample is drawn, that
true true population paramet true population param
intervalinterval isinterval is estimatedinterval is estimated the
is either in or not in the interval.

An analogy can be made to flipping a coin.

BeforeBefore aBefore a fair coin is flipped, there is a probab
0.50.5 that the coin will0.5 that the coin will display heads.0.5
associatedassociated with the *random event* of flipping the of
OnceOnce the random event hasOnce the random event has oc
of confidence that the coin will display heads.

StatedStated another way, if sampleStated another way, if sa
repeatedly drawn from the same drawn from the same popula
confidenceconfidence intervconfidence interval wconfidence
thethenthen the relative frequency of those intervals ththe
containcontain the true unknown population pcontain th
would *approach* 95 percent.

This is a *frequency*-based view of probability

ThisThis is seen in the hypotheticalThis is seen in the hypothe
Figure 4.

TheThe hypothesized mean of zeroThe hypothesized mean of
populationpopulation parameter) is containedpopulation par
ofof of the intervals; 48 out of the 50 confidence intervals,
oror 96% of the time, the intervaor 96% of the time, the in
population parameter.

ThTheThe The nuThe null hypothesis of the peach investigatio
be evaluated by constructing a *confidence interval*.

IfIf we let represent the *significance level,* then the
100(1100(1)% *confidence interval* will yie will yield will
results.

SinceSince = 0.10, a 100 = 0.10, a 100(1)) or 90% *confid*
using the 10 sample differences is constructed.

A 90% confidence interval for theA 90% confidence interval f

$$
\overline{} \qquad\qquad \left(\frac{}{\sqrt{}} \right),
$$

where,

- $\overline{}$ is the parameter estimate,

- $t_{(9,(9,(1\ \ 0.10/2))}$ is is the tabled is the tabled value for a *2-sided co* *interval* using a *t*-distribution with (10 1) = 1) = 1) = degreesdegrees of freedomdegrees of freedom and a confi (1 0.10) = 0.90; and

- $\left(\frac{}{\sqrt{}} \right)$ is the standard deviation of the sample differencesdifferences divided by the squadifferences samplesample size, the is referred tsample size, the is error of the mean differences, $\overline{}$.

Why use $t_{(9,\ (1 \quad 0.10/2))}$?

" FForFor a 0.90 *confidence coefficient*, 10% of the area, 10% outside of the *confidence interval*.

" SinceSince this is a two-siSince this is a two-sideSi distributiondistribution is symmetric,distribution is sym the UCL and 5% of the area is below the LCL.

" $t_{(9,(9,\ (1 \quad 0.10/2))}$ is used rather than $t_{(9,\ 0.10/2)}$, in order to have a positive *t*-value. Note that,

$$t_{(9,\ (1 \quad 0.10/2))} = |\, t_{(9,\ 0.10/2)} \,|,$$

since the *t*-distribution is symmetric.

TheThe t-value that has 5 percent of the-value that has 5 pe
curve to its right is $t_{(9,\ (1 \quad 0.10/2))}$ = 1.83.

ForFor the peach investigation (usinFor the peach inves
statistics on page 66),

$$— \quad t_{(9,\ (1\ \ 0.10/2))} \left(\frac{}{\sqrt{}} \right)$$

$$= \quad \left(\frac{}{\sqrt{}} \right)$$

$$= \ 4.39 \pm 5.38$$

$$= \ [\ \ 1.00,\ 9.77].$$

TTheThe 90% *confidence interval* for the mean difference for t
[1.00, 9.77].

SinceSince the hypothesized mean, $_{Do}$ = = 0, falls = 0, falls wi
90%90% *cconfidence confidence interval*,, , there, there , there is, th
to reject the H_o at the 0.10 *significance level*.

For evaluating hypotheses of the form,

$$H_o: \quad = \quad _o \textit{ versus } H_a: \quad\quad _o,$$

thethe 100(1)% *confidence inteconfidence intervaconfidenc* equivalentequivalent to a *two-sided* hypothesis test of the mean the *level of significance*.

Interval Estimation ***Example One-Sided***

WeWe can calculateWe can calculate a *coconfidence limit* t directional hypotheses,

H_o: $_D$ $_{Do}$ *versus* H_a: $_D$ > $_{Do}$.

ForFor a *one-sided alternative* in t in the ri in the right distribution,distribution, we need to calcudistribution, we r *limit*,

LCL = {sample estimate margin of error}.

IfIf the hypothesized value, $_{Do}$, , is less than the LCL thenthen conclude therethen conclude there isthen conclude t null hypothesis.

Interval Estimation *Example One-Sided*

ForFor the peach investigation (usinFor the peach inves statistics on page 66),

$$\text{LCL} = \overline{} \quad t_{(9,\,(1 \quad 0.10))}\left(\frac{}{\sqrt{}} \right)$$

$$= \quad \left(\frac{}{\sqrt{}} \right)$$

$$= 4.39 \quad 4.06$$

$$= 0.33.$$

TheThe 90% *lowerlower confidence limit* for the for the mean di is 0.33.

SinceSince the hypothesized mean,Since the hypothesized me LCLLCL =LCL = 0.33, there is enough evidence toLCL = 0.33 at the 0.10 *significance level*.

Interval Estimation ***Example One-Sided***

Why use $t_{(9,(1\text{-}0.10))}$?

" FForFor a 0.90 *confidence coefficient*, 10% of the area, 10%
outside of the *confidence interval*.

" SinceSince this is a *one-sidedone-sided limit* and the alternat
inin the right tail, all 10% oin the right tail, all 10% of t
LCL.

" SinceSince the *t*-dis-distribution-distribution -distribution
coefficient is used rather than th is used rather than the is use
order that the *t*-value is positive.

For evaluating hypotheses of the form,

H$_o$: $_o$ *versus* H$_a$: > $_o$,

thethe 100(1)% *lower confidence limit* for the for the mean
equivalentequivalent toequivalent to a *one-sided* hypothesis tes
the *level of significance*.

Interval Estimation & Hypothesis Tests *References*

(1)(1) Moore,Moore, David S., (1997). *Statistics Concepts andStatistics Concepts and Controversies*,, 4th edit
W. H. Freeman and Company.

AA very well written book on statistA very well written book on statistical concepA very we
understandunderstand data in the face of uncerunderstand data in the face of uncertainty,
exploreexplore the data, and how toexplore the data, and how to draw conclusions from data. T
fewfew equations, but rather the focus is on the thought process behindfew equations, but r
arguments and how to be a critical consumer of scientific studies and surveys.

(2) Utts, Jessica M., (1996). *Seeing Through Statistics*, Belmont CA: Duxbury Press.

AnotherAnother very well written book onAnother very well written book on statistical concept
toto help the readerto help the reader understand scientific studies and surveys,to help the read
andand accurate information from useless and misleading; and to interpand accurate informat
inferentialinferential methods, by focusing on how to understand them rather than how to
computecompute them. Utts uses a case-study appcompute them. Utts uses a case-study
interpretation of data that appear in the media and scientific literature.

(3) Kanji, Gopal K. (1997). *100 Statistical Tests*, SAGE Publications, London.

ThisThis is a great referenceThis is a great reference that begins with aThis is a great reference tha
TheThe remainder of the book presents the vaThe remainder of the book presents the va
parametricparametric and non-parametric. Each statistical test iparametric and non-parametr
test,test, the limitations, how to conduct thetest, the limitations, how to conduct the ttest, the
example. This is a very useful reference.

(4) Ott,Ott, Lyman (1988). *An IntrodAn Introduction to Statistical MeAn Introduction to Statistica*
edition, PWS-Kent Publishing Company, Boston.

ThisThis book is appropriate for someone with little or no exposure to statistical
methodsmethods and data analysis. methods and data analysis. Itmethods and data analysis. It a
prior knowledge of statistics.